Journal of Applied Geophysics 150 (2018) 52–66

Contents lists available at ScienceDirect

Journal of Applied Geophysics

journal homepage: www.elsevier.com/locate/jappgeo

Inferring the most probable maps of underground utilities using Bayesian mapping model

Muhammad Bilal [a,*], Wasiq Khan [f], Jennifer Muggleton [c], Emiliano Rustighi [c], Hugo Jenks [d], Steve R. Pennock [d], Phil R. Atkins [e], Anthony Cohn [b]

[a] University of California, Center for Environmental Implications of Nanotechnology (UCCEIN), Los Angeles, USA
[b] School of Computing, University of Leeds, Leeds, United Kingdom
[c] Institute of Sound and Vibration Research, University of Southampton, Southampton, United Kingdom
[d] School of Electronic and Electric Engineering, University of Bath, Bath, UK
[e] School of Electronic, Electrical and Computing Engineering, University of Birmingham, Birmingham, United Kingdom
[f] Manchester Metropolitan University, Manchester, United Kingdom

ARTICLE INFO

Article history:
Received 13 February 2017
Received in revised form 23 December 2017
Accepted 11 January 2018
Available online 31 January 2018

Keywords:
MTU sensors
Most probable maps
Bayesian data fusion
Image processing
Bayesian regression

ABSTRACT

Mapping the Underworld (MTU), a major initiative in the UK, is focused on addressing social, environmental and economic consequences raised from the inability to locate buried underground utilities (such as pipes and cables) by developing a multi-sensor mobile device. The aim of MTU device is to locate different types of buried assets in real time with the use of automated data processing techniques and statutory records. The statutory records, even though typically being inaccurate and incomplete, provide useful prior information on what is buried under the ground and where. However, the integration of information from multiple sensors (raw data) with these qualitative maps and their visualization is challenging and requires the implementation of robust machine learning/data fusion approaches. An approach for automated creation of revised maps was developed as a Bayesian Mapping model in this paper by integrating the knowledge extracted from sensors raw data and available statutory records. The combination of statutory records with the hypotheses from sensors was for initial estimation of what might be found underground and roughly where. The maps were (re)constructed using automated image segmentation techniques for hypotheses extraction and Bayesian classification techniques for segment-manhole connections. The model consisting of image segmentation algorithm and various Bayesian classification techniques (segment recognition and expectation maximization (EM) algorithm) provided robust performance on various simulated as well as real sites in terms of predicting linear/non-linear segments and constructing refined 2D/3D maps.

1. Introduction

The costs associated with street works in the UK is of critical consideration due to the vast majority of utilities buried underneath the roads and their repair/(re)installation (£7b annual) (Mcmahon et al., 2005). The types of utilities buried under the ground are diverse and their amount is notoriously large which makes excavation a challenging task in order to upgrade these underground networks. In addition, the statutory records of underground networks are typically incomplete and inaccurate particularly for old street works (Burtwell and E. A., 2004). An important undertaking is to develop schemes to detect what is buried underground that could be associated to their records and could become cost savior. A multi-sensor mobile laboratory MTU (Underworld, 2011) was developed which consists of multiple sensors capable of deploying several approaches to detect different types of buried infrastructure. The MTU device, was designed to assess the feasibility of a range of potential technologies that can be combined into a single device to accurately locate buried pipes and cables. The potential technologies included ground penetrating radar (GPR), low-frequency quasi-static electromagnetic fields (LFEM), passive magnetic fields (PMF) and low frequency vibro-acoustics (VA) and significant advances have already been made (Royal et al., 2011; Royal Acd et al., 2010).

The location estimation approaches combined by MTU provide significant advantages over other commercially available techniques (Ashdown, n.d.) for detecting wide variety of utilities and control trials were taken for test commercial sites. As a result, excavations necessary for maintenance and repair can be largely reduced using such device. An important undertaking is to use heterogeneous information from these sensors and build refined maps of buried utilities in real time. However, due to the heterogeneity in features of utilities and ground properties, it

* Corresponding author.
E-mail address: m.bilal@ucla.edu (M. Bilal).

https://doi.org/10.1016/j.jappgeo.2018.01.006

is challenging to develop a general technique that could assess heterogeneous information and handle the uncertainties associated to this task. The integration of information obtained from multiple sensors on MTU is of critical importance in order to make sense of the data before providing a precise information on a site. The knowledge obtained from different sensors presents itself non-symbolically i.e. the delivered data is essentially an image representing what the sensor "sees" underground. In contrast, utility records are almost universally represented symbolically i.e. they are stored in a spatial database as records with a vectorized representation of their spatial position, along with attribute information (such as material, diameter). It is therefore challenging to provide a useful and accurate representation of the data acquired from a variety of sensors. Therefore, a data fusion approach consisting of automated techniques for data extraction and integration was imperative.

The map (re)construction model developed in this work was an improvement over (Chen and Cohn, 2011) which was initially designed only for 2D construction of the map assuming that it consists of only linear segments. In addition, the data preprocessing for hypotheses extraction in (Chen and Cohn, 2011) was not combined as a complete model and it was assumed that the hypotheses were extracted from GPR images using an iterative clustering/classification techniques prior to data fusion tasks. Simple clustering/classification algorithms for hypotheses extraction such as k-means or Dbscan were restricted in several ways for asset classification problem when developing real time maps. For example, traditional k-means clustering algorithm creates the clusters based on Euclidean distance of each data point to the centroids (initially selected randomly). Also, the number of clusters to be created is known in k-means algorithm. Depending on statutory records to identify the number of segments was not reliable as, even providing valuable information, they are inaccurate and may contain incomplete information. Dbscan (Sander et al., 1998) also separates the clusters based on Euclidean distance without providing the desired number of clusters to be generated as prior. However, Dbscan requires radius in order to differentiate the clusters that is used as a criterion for decision making on number of clusters. The Euclidean distance between parameters is important in both approaches which is helpful in situations where clustering is only distance based.

Bayesian data fusion models have been utilized for numerous applications and there is a large body of literature proposing Bayesian modeling for data fusion and uncertainty management, thus, providing motivation for the work proposed in this study. To date, Bayesian modeling has been successfully implemented in similar applications, such as seismic/Magnetotelluric inversion (Dettmer et al., 2014; Guo et al., 2011), water distribution management, modeling for rock-physics analysis, gas and buried near-surface utility mapping (Ristić et al., 2017; Ji et al., 2016; Wang and Lu, 2016; Ren et al., 2017; Aleardi et al., 2017; Fernández-Martínez et al., 2013). Among several impactful studies using Bayesian modeling, the approach of combining multiple data sources and Bayesian data fusion for bedrock tracking has been of significant interest such as (Fiannacca et al., 2017; Christensen et al., 2015; Oldenborger et al., 2016). These studies proposed automated tracking of bedrock depth and orientation by combining data from different inversion models, borehole data (Christensen et al., 2015), and the utilization of time-domain electromagnetic data (Oldenborger et al., 2016) to systematically handle uncertainties in data of heterogeneous nature and reconstruct estimated maps of bedrocks. An application of Bayesian data fusion approach for the prediction of water pipe failures was developed by (Oldenborger et al., 2016) with the capability to be integrated with the geographical information system of water resources and automatically predicting pipes of potential failures. Another application of neural networks and pattern recognition was developed utilizing only ground penetrating radar (GPR) data (images) to train the model on hyperbolic features (of buried objects) and predict the locations and depths of buried solid objects followed by automatic construction of the maps of underground solid objects (pipes and cables) (Ristić et al., 2017; Al-Nuaimy et al., 2000). It is noted that, in addition to the inclusion of GPR image analysis as proposed by (Al-Nuaimy et al., 2000), the work proposed in this paper provides wider applicability due to the inclusion of multiple sensors of the MTU device and the application of Bayesian models being capable of incorporating incremental learning (unlike neural networks) upon the acquirement of new knowledge.

In other similar works, Neira (Neira and Tardos, 2001) developed a data association model for addressing the problem of robust data association for simultaneous vehicle localization and map building which was an improvement over gated nearest neighbor (NN) (Bar-Shalom, 1987) for tracking problems that successfully rejects spurious matching and provides optimal solutions in terms of pairs of matching in cluttered environments. The correlation between measurement prediction errors in 2D space in cluttered environment provides robust data association with an efficient traversal of the solution space. However, the directional errors (linearity) caused mismatching of the segments with manholes using the hypotheses extracted from the sensors. Abhir and Roland (Bhalerao and Wilson, 2001) also used a Multi-resolution Fourier Transform (MFT) for capturing sufficient shape and orientation of objects within a given image. The use of statistical analysis and camera projections to estimate the location/orientations of line segments in 3D image was also implemented for similar linear segment construction problems (Dong-Min and Dong-Chul, 2009; Chen and Wang, 2010). However, these approaches are only limited to an image of objects and segments which is used to reconstruct a 3D image. MTU mapping, on the other hand, is multi-source data fusion approach to integrate information from multiple sources and produce most probable maps utilizing advanced machine learning/data mining techniques. For linear segment fitting, significant amount of literature report the use of different regression models including EM algorithm that can efficiently fit at higher accuracy levels (Ward et al., 2009; Ester et al., 1996; Sanquer et al., 2011; Delicado and Smrekar, 2007; Werman and Keren, 1999; Friedman and Popescu, 2004). The classification of data samples based on its source as distinguished by MTU sensors is, however, lacking in these approaches as these algorithms were developed for regression scenarios. In addition, the connection establishment (manhole-segment) was not considered as an underlying issue as only the general regression was covered.

The Bayesian mapping model is capable of using automated techniques for hypotheses extraction, classification, segment recognition and connection establishment with the associated manholes. We associate a probability distribution with every such hypothesis reflecting possible errors in the measurements (uncertainty due to the fusion of data from multiple sources) and hypothesis extraction process. These geographical positions ($\boldsymbol{x}, \boldsymbol{y}$) and depths ($\boldsymbol{z}$) were used as input to the next stage of the mapping system. A variety of Artificial Intelligence (AI) techniques and algorithms were implemented such as Bayesian Data Fusion (BDF), image segmentation, orthogonal distance hyperbolic fitting, and weighted variation. The algorithms for automated data processing and map (re)construction were developed for real time operative capability of MTU device. A complete use case can be tested using real time mapping model where hypotheses extraction techniques were combined with iterative connection establishment and visualization techniques. Several simulated as well as real sites were tested, and it was demonstrated that the model is robust in various conditions where statutory records were unavailable, and the sensor readings were sparse. The segments were recognized and noise was removed successfully in various situations for mapping the utilities demonstrating the ability of model to work in real time complex situations.

2. Materials and methods

The model for Bayesian mapping followed the workflow depicted in Fig. 1. The sequential steps in model workflow were as follows; (1) data preprocessing, (2) segment recognition, and (3) segment-manhole

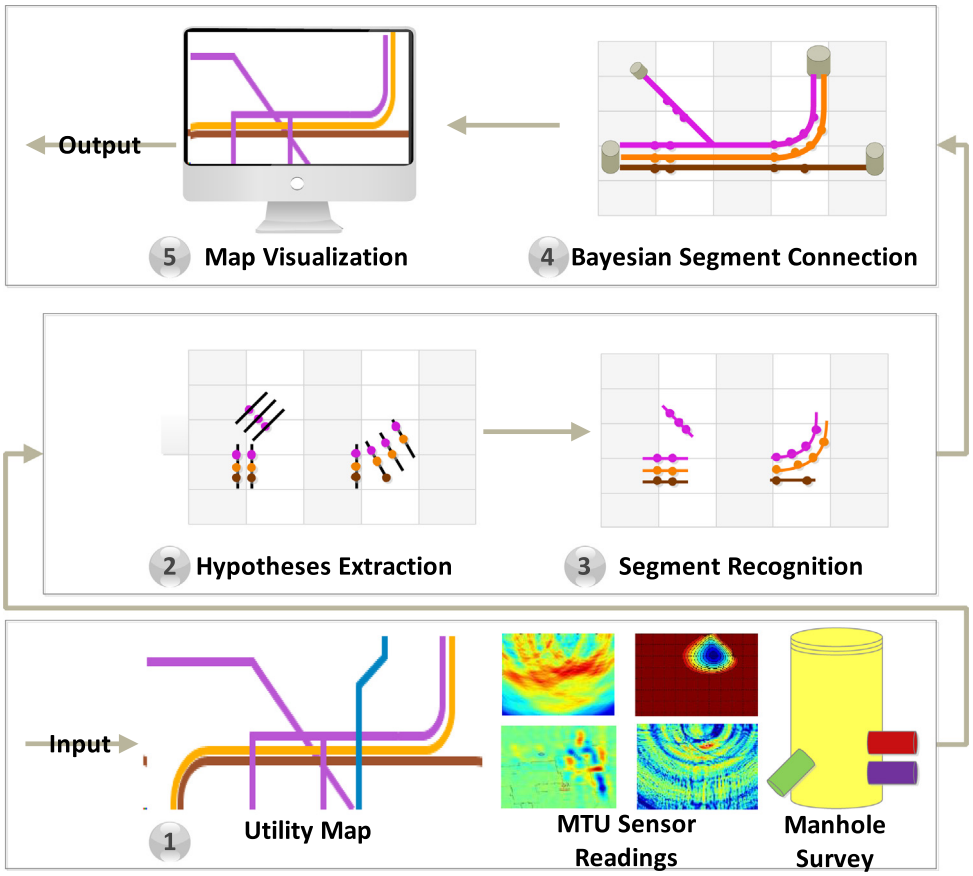

Fig. 1. Model Workflow (Data preprocessing for hypotheses extraction following segment recognition, Bayesian segment connection and refined map visualization.

connection. The datasets consisted of both qualitative (raw images from MTU sensor device, statutory records (utility maps)) as well as quantitative information (manhole surveys providing information on directions and depths). More details are given in Section 2.1.2 (MTU sensor raw data). The data from these sources were integrated together in the form of positions (x, y), depths (z) and orientations (Θ) as an

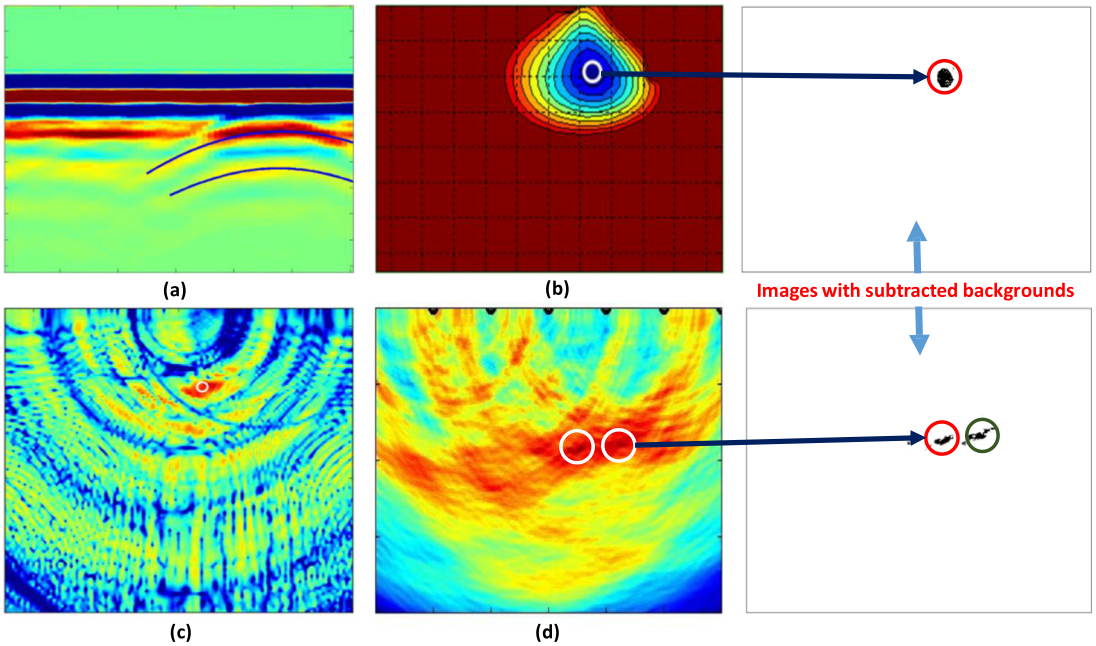

Fig. 2. Images taken from MTU sensors and processed to mark the hypotheses (hyperbolae and circles). The images were created from the raw data collected from **(a)** Commercial GPR, **(b)** Passive Magnetic Field, **(c)** MTU GPR, and **(d)** Vibro acoustics. (Right) Images **(b)** and **(d)** with backgrounds subtracted using image segmentation technique to extract hypotheses.

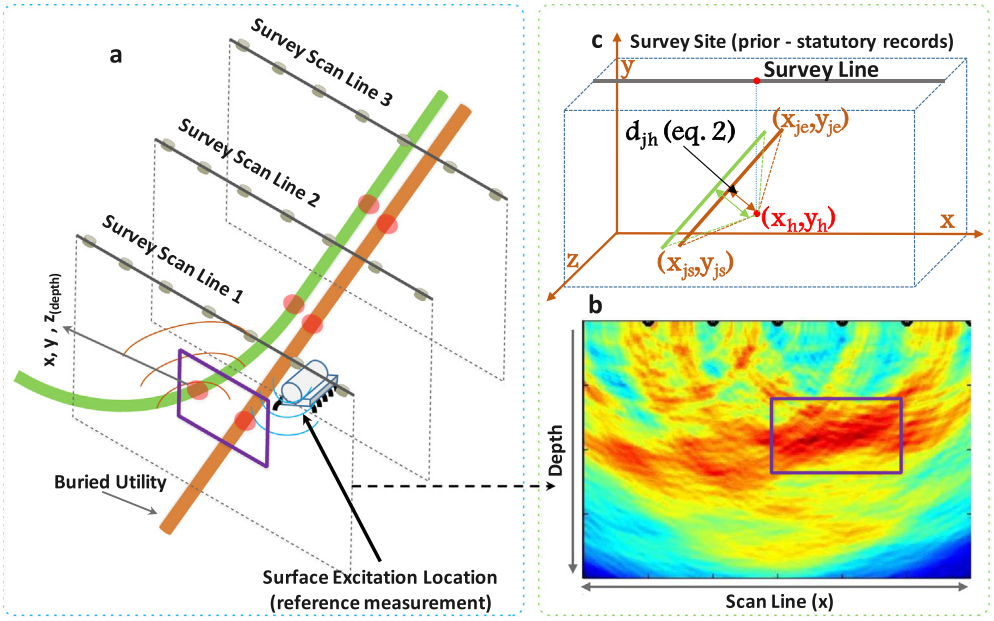

Fig. 3. (**a**) MTU scan lines on the surface to record reflections w.r.t. surface excitation and reference measurement location. (**b**) Correlated functions to create raw images using frequency domain transformation, and (**c**) hypotheses extracted from raw images using adaptive image segmentation techniques, and integrated with utility records for segment recognition algorithm.

input for segment recognition algorithm. The geometrical measurements (i.e., readings of slope distances in horizontal and vertical planes) with respect to spatial survey structure were acquired using the Total Station Theodolite (TST). The relative positions of the MTU mobile laboratory were, therefore, called the survey lines at given (***x, y***) coordinates from the TST. Manhole surveys were helpful in determining the types of buried utilities at a site and estimating their orientations and depths. Bayesian recursive algorithm as known as EM was developed for segment-manhole connection. EM is an iterative algorithm for finding the maximum-likelihood estimation of a set of parameters for a specific distribution in a statistical model when it contains unobserved latent variables (Do and Batzoglou, 2008). Based on partially connected information from different sources and segments recognized from segment recognition, EM algorithm was used to identify most probable connections of segments with assets and connect them in an iterative manner. The depth information from manholes and sensor readings were then used in conjunction with EM output to produce refined 2D/3D maps of an investigated site.

2.1. Data preprocessing

Data for model development were collected from following sources;

2.1.1. Statutory records

The utility/statutory records consisted of maps of buried assets as ground truth which provided information on approximate orientations

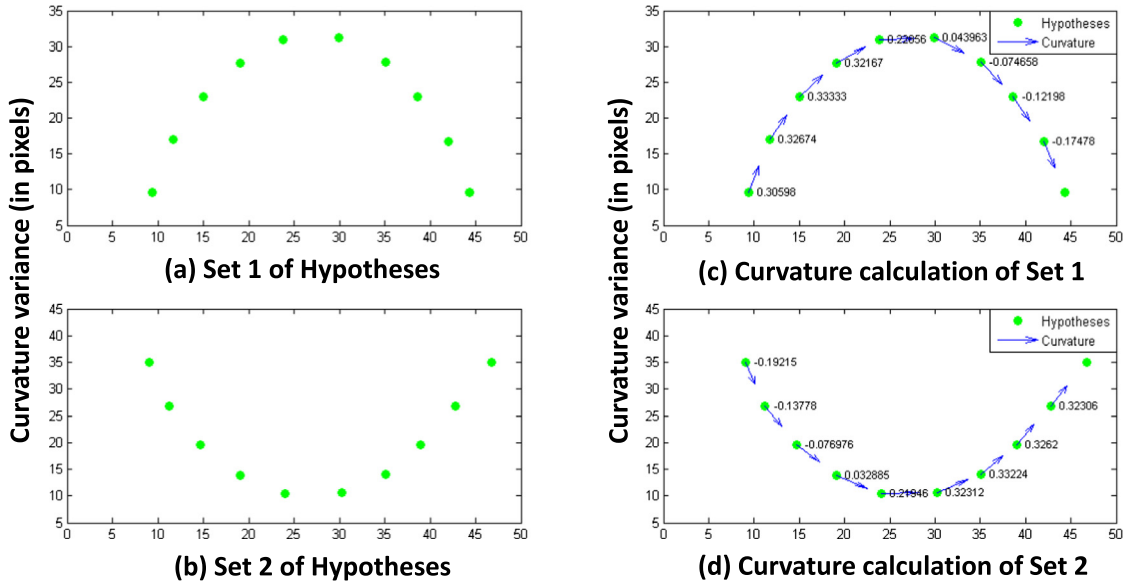

Fig. 4. Clusters with curvature of hyperbolic features.

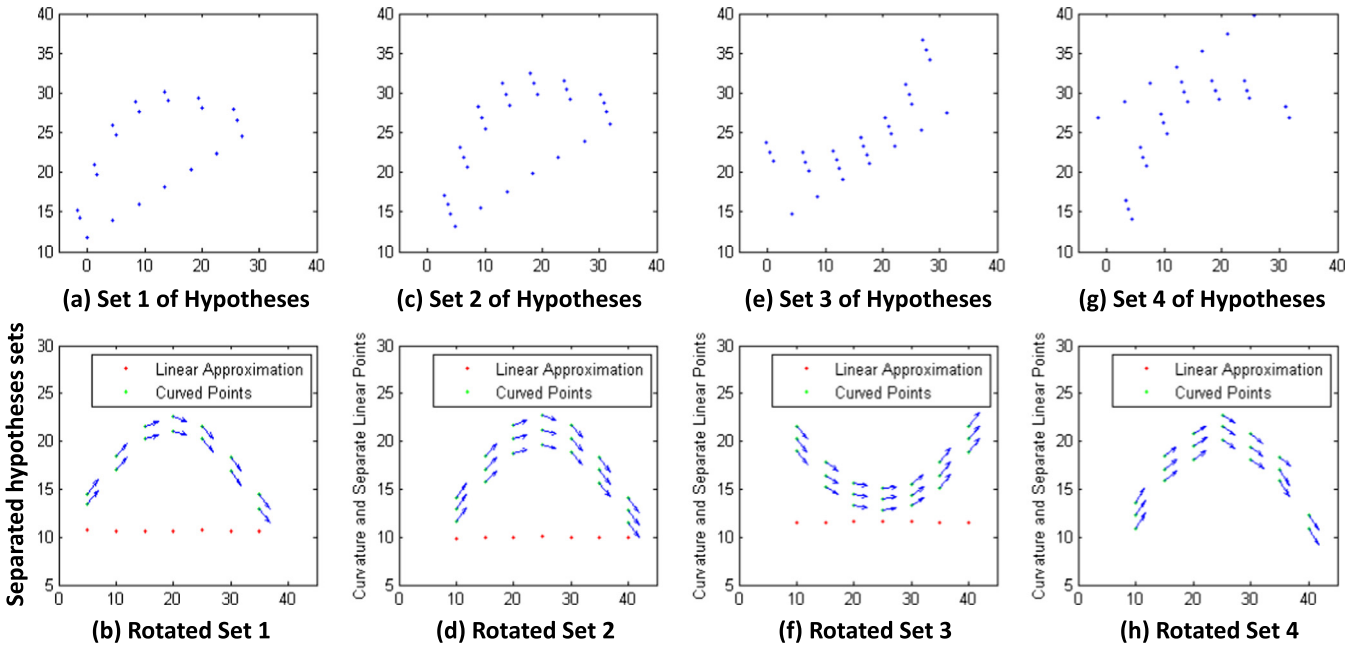

Fig. 5. Separation of hypotheses into cluster of linear clusters and curved clusters. For clusters with circular features, a three-point circle fitting was used, as in Fig. A-1 (Appendix).

and positions of these assets. Since these records are inaccurate, their position information was not included in segment recognition algorithm. However, the information pertaining to the type of buried asset and their approximate directions was helpful in segment-asset connection algorithm. Therefore, the positions and orientations $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\Theta})^{T}$ from these records were used for segment-manhole connection establishment.

2.1.2. MTU sensor raw data

The raw data included heatmaps (Fig. 2) of investigated sites that were produced from various MTU and commercial sensors. The sensors used in MTU mobile apparatus include Ground Penetrating Radar (GPR), Passive Magnetic Field (PMF), Vibro Acoustic, and Low Frequency Electro-Magnetic (LFEM). Briefly, the sensors and their functionality are explained in this section. Detailed information on each sensor can be found elsewhere (Dutta et al., 2013; Muggleton et al., 2011; Thomas et al., 2009).

2.1.2.1. Ground penetrating radar (GPR). GPR locates buried utilities (metallic and non-metallic) by transmitting electromagnetic waves into the ground and collecting the response (waves reflected from the objects underground) (Ristić et al., 2017; Al-Nuaimy et al., 2000). A GPR scan produces an image as a collection of multiple A-scans of reflected waves received at different wave-travel times and integrating them into a B-scan. In B-scan, the rows represent depth (by utilizing the reflected waves) and columns represent horizontal positions of scan lines.

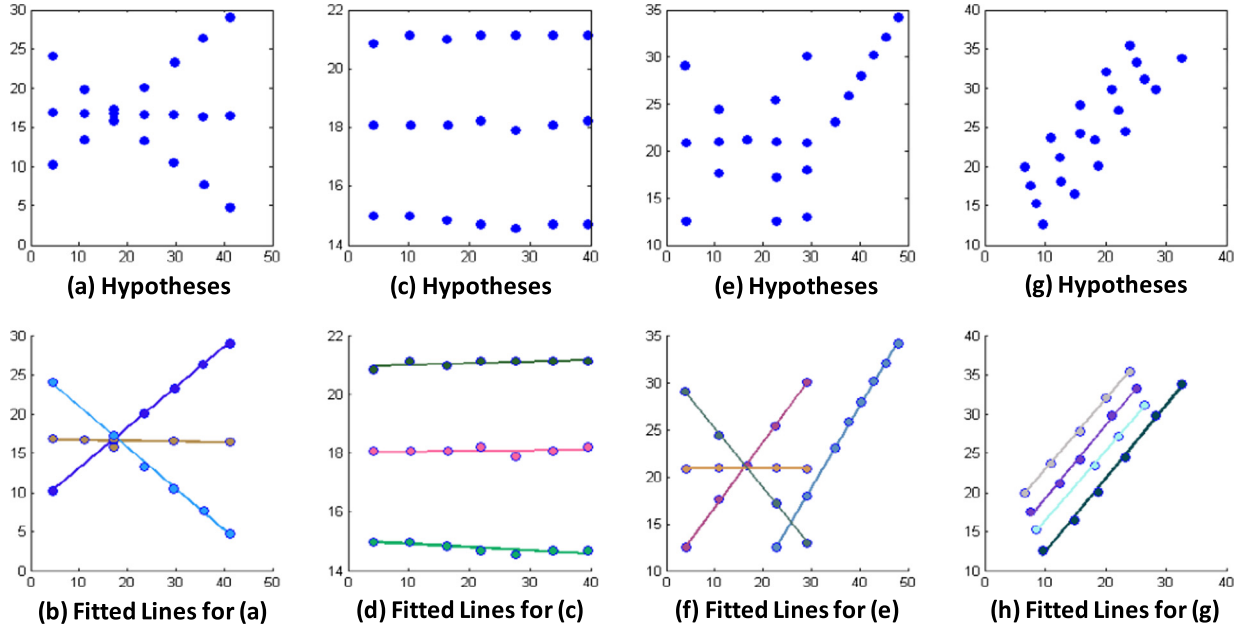

Fig. 6. Segment recognition algorithm fitting linear segments in groups of hypotheses.

Table 1
Settings for simulated and real sites (S# = simulated site #, R# = real site #).

Data	# Manholes	# Pipes & cables	# Sensor readings
S1	10	13	28
S2	12	16	30
S3	26	46	87
R1	19	8	18
R2	8	12	25

2.1.2.2. Passive magnetic field (PMF). PMF is useful for detecting the cables by utilizing the magnetic field generated by current in a cable as well as neighboring objects which generate magnetic field (may also be from current flow leak from a buried cable). The cluster centroids (Fig. 2c) are multiplied by the surveyed map dimensions ($m \times n$ image) which are then divided by surveyed lengths and depths, respectively, to calculate the ($\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$) coordinates on the map.

2.1.2.3. Vibro acoustic (VA) sensors. VA device at the MTU uses seven geophones for recording the reflected velocities of the waves that are generated by exciting the surface. The response received at the surface is utilized to create a cross-sectional image with respect to the reference measurement used near the exciting location. The cross-sectional image is created via the time-domain transformation of the correlation functions between geophone and reference measurement recordings (Figs. 2d, 3). The images are then analyzed to extract hypotheses using adaptive image segmentation techniques (Section 2.1.4, Hypotheses extraction).

2.1.2.4. Low frequency electromagnetic field (LFEM). LFEM approach is based on injecting the current into the ground and measuring the voltage on the surface via coupled plates moved along the surface. The LFEM approach at MTU is operated in a targeted grid location and the resulting image represents the underground structure.

Using the above approaches, qualitative raw images consisting of $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})^{T}$ information (x-axis as the length of scan line and y-axis as the depth of the investigated surface (typically 2–4 m deep)) were collected and processed to extract hypotheses which were utilized for 2D/3D map reconstruction. The orientation of MTU sensor device provides approximate direction of buried asset as the survey is usually taken in the direction perpendicular to the orientation of asset.

2.1.3. Manhole survey

Manhole surveys for real sites were conducted to collect supporting information on estimated orientations and depths of buried assets. In addition, the types of buried assets were also recorded to distinguish between them when constructing site maps. The quantitative information from manhole surveys therefore included $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\Theta})^{T}$ as well as the type of buried asset as qualitative information.

2.1.4. Hypotheses extraction

Hypotheses (processed coordinates of the location detected by MTU sensors) extraction was implemented on raw sensor image data to automatically estimate the positions from surveyed locations. The images were segmented into clusters of foreground (hypotheses) and background pixels using unsupervised image segmentation techniques. The adaptive image segmentation was used to highlight and quantify

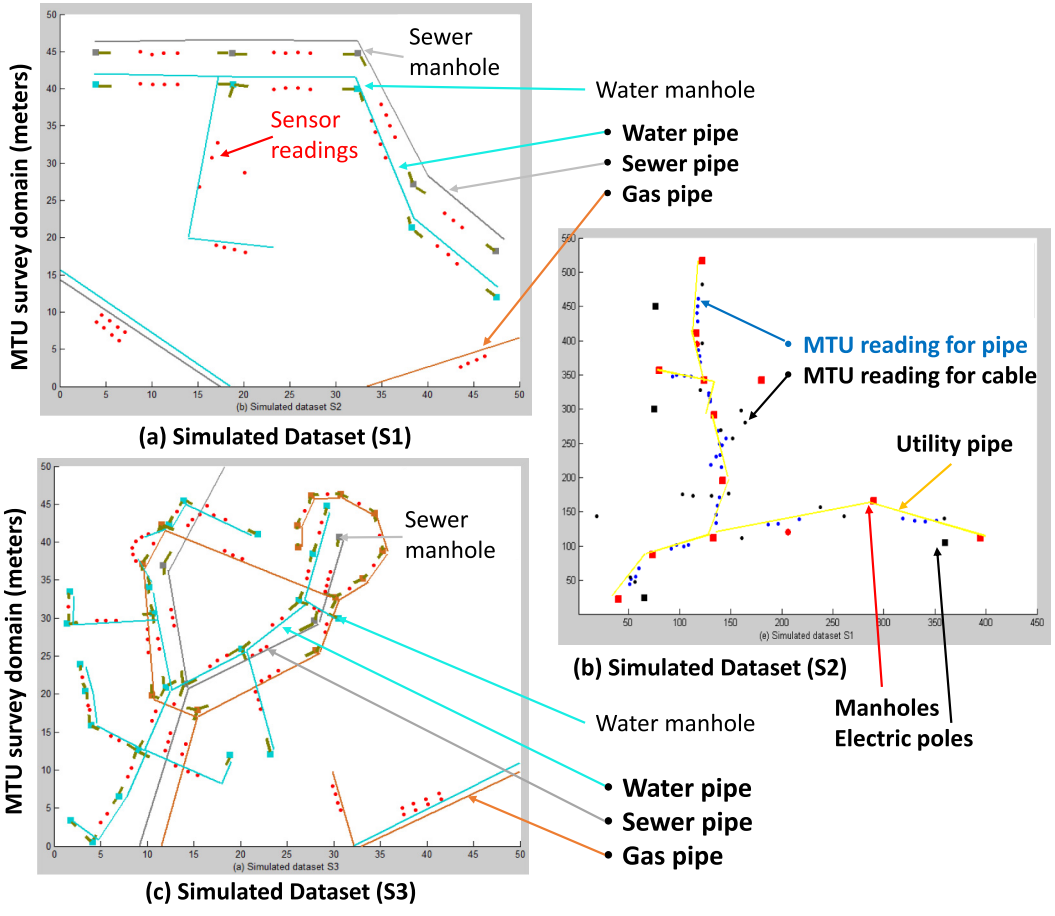

Fig. 7. Simulated datasets with hypotheses points and initial maps for **(a)** S1, **(b)** S2, and **(c)** S3. Squares correspond to manholes (legend) and circles correspond to sensor readings.

regions of high thresholds in the image (Nobuyuki Otsu, 1979). Images were initially enhanced based on histogram equalization in order to increase image contrast (Krishna, 2013; Gonzalez, 2008). Subsequently, the adaptive segmentation algorithm (Nobuyuki Otsu, 1979), which regards an image as a data clustering scenario, was applied to divide the image into clusters of two classes (foreground and background) which iteratively finds the threshold that minimizes the weighted within-class variance and thus, maximizing the between-class variance. The centroids of clusters were identified as approximate positions of assets (e.g., Fig. 2**(b and d)** and their backgrounds subtracted (Fig. 2 **(right)**)). Sensor output image is often noisy and may not provide an accurate location. Additionally, multiple hypotheses locations may be reported in close proximity when different types of assets (or a non-asset object) are present as given in Fig. 1 (hypotheses extraction). It is noted that the sensory scan lines may consist of multiple hypotheses of different types or a single hypothesis given the ground truth (i.e., utility map in Fig. 1). Accordingly, the image segmentation algorithm provides clusters (regions) of possibly one or more locations given (i) the type of buried material from utility map, and (ii) raw images provided by different MTU sensor. From multiple images of a single surveyed location, the most probable locations of assets were estimated using Bayesian weighted average technique, i.e.

$$h(x_i) \approx N(x, \mu, \Lambda) \tag{1}$$

In the above equation, x_i denotes the relative location of hypothesis i (w.r.t. MTU scan line) collected from the image, μ is the mean of selected hypotheses, and $\mathbf{\Lambda}$ is covariance matrix of the hypotheses. Maximum A Posteriori (MAP) position of an asset was obtained using $h(x) = \arg\max \prod_{i=1}^{n}(h(x_i))$ where n is the number of hypotheses.

Prior to segment recognition from groups of hypotheses, the identification of noise was conducted in terms of points with no manhole connection within a given previously tested threshold ($\delta = 2$ m). Each hypothesis point was validated using: (1) point to line (nearest buried utility in statutory record) distance and (2) manhole orientations. The distance d_{jh} (Fig. 3c top right) from each pipe segment j to the point h (Fig. 2d (white circle, Fig. 3b) was calculated using

$$d_{jh} = \frac{\left|(x_{je}-x_{js})\left(y_{js}-y_h\right)-(x_{js}-x_h)\left(y_{je}-y_{js}\right)\right|}{\sqrt{(x_{je}-x_{js})^2+\left(y_{je}-y_{js}\right)^2}} \tag{2}$$

where (x_{js}, y_{js}) and (x_{je}, y_{je}) are the start and end points of the pipe segment j respectively, and (x_h, y_h) is the location of hypothesis point. The point h was considered an orphan point if its distance from a pipe segment j was greater than δ. i.e. $h = \begin{Bmatrix} orphan & if\ d_{hk1}>\delta\ and\ d_{hk2}>\delta \\ h_r & otherwise \end{Bmatrix}$ where $d^s_{h,k}$ and $d^e_{h,k}$ are the distances between of k manhole (with position (x_k, y_k), where the segment of same type exists (i.e., a pipe connection matching pipe segment j in this case)) from the segment start and segment end, respectively. The probability that the point h belongs to a segment j was then calculated as $S_j = \arg\max_j(P_{hj})$ where $P_{hj} = F(h \to j | d_{jh}) \approx N(h, d_{jh.})$

2.2. Segment recognition

The hypotheses obtained from above step were given as input to segment recognition algorithm which estimated the segments by joining hypotheses together based on their positions and orientations. The input to segment identification algorithm included the hypotheses and the orientations in which the surveys were taken (usually from left-to-right). The linear segments were iteratively classified from the groups of hypotheses which were located within the range of an angle of up to $\varepsilon = \pm 0.08$ (i.e., wider sensitivity of up to 4.5 degrees). The algorithm starts with first two chosen hypotheses as assigned to first class (starts with class = 1), and the orientation from the first hypothesis to all other hypotheses are calculated. The next point is chosen and assigned to the same class if their difference in orientations is $\leq \varepsilon$ otherwise the chosen point is assigned a new class. The algorithm is repeated for each hypothesis and provides the output as points assigned to a class which are then joined using least squares, circle and/or polynomial fitting techniques. The workflow of the algorithm is given in Algorithm 1.

Algorithm 1. Segment recognition algorithm (Pseudo code).

```
Initialize first hypothesis point to class 1 (h_1 = 1)
  while: no hypothesis point is not assigned a class
    i = i+1
    if: angle between h_1 and h_i is higher than threshold
      for: j ← 1st to i^th hypothesis point
          if: angle between h_j is h_i is lower than threshold
              assign class C_j to h_i
          else:
              assign new class to h_i
      end
    else:
        if: h_j matches an existing class
            Assign existing class
        else:
            assign new class to h_i
    end
  end while
Group hypotheses of same classes into clusters
```

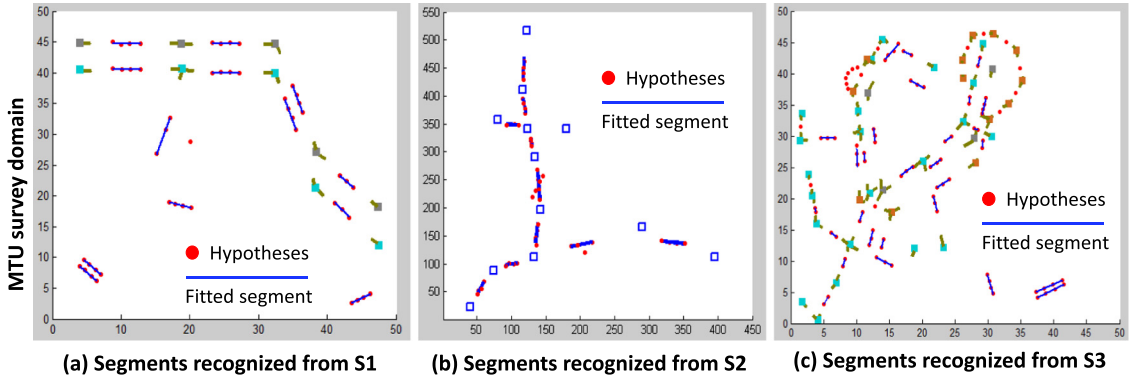

Fig. 8. Segments recognized for **(a) S1**, **(b) S2** and **(c) S3**. Squares in the figures correspond to different manholes (legend) and red circles correspond to sensor readings. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

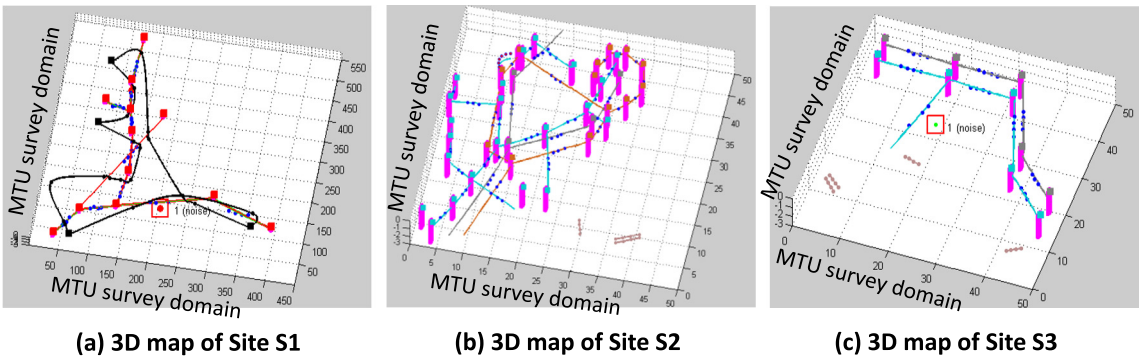

Fig. 9. 3D map construction for segment-manhole connection (S1(**a**), S2(**b**) and S3(**c**)).

To separate the clusters of linear segments from curved segments, the goodness of fit (R^2) was checked and compared to threshold $\varepsilon_1 \geq 0.99$ (0.99 confidence). Each cluster with $R^2 \geq 0.99$ was separated as a linear pipe segment and the remaining clusters were considered as curved or circular segments. For each hypothesis point in curvature analysis, the tip of the vector $\boldsymbol{r(t)} = <\boldsymbol{x(t)}, \boldsymbol{y(t)}>$ traces out a path in the plane where t represented MTU survey line and y represented the change in positions represented by hypotheses. The relative positions of the hypotheses points were therefore represented as follows $\boldsymbol{r'(t)} = <\boldsymbol{x'(t)}, \boldsymbol{y'(t)}>$ where $\mathbf{x}'(t) = \frac{\partial x}{\partial t}$ and $\mathbf{y}'(t) = \frac{\partial y}{\partial t}$. The unit tangent vector was calculated as $\mathbf{T} = \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}$, where $\|\mathbf{r}'(\mathbf{t})\| = \sqrt{\mathbf{x}'(\mathbf{t})^2 + \mathbf{y}'(\mathbf{t})^2}$. In order to determine the variation in the position ($\boldsymbol{y}$) with respect to ($\boldsymbol{x}$), the difference interval was obtained using $\eta = \| \arg\max_{i \in N}(\mathbf{T}_i) - \arg\min_{i \in N}(\mathbf{T}_i) \|$. Testing data clusters of various curvatures led to the selection of approximate curvature threshold to 0.15. The segments were assigned as curved using

$$Curved = \begin{cases} true \\ false \end{cases} \quad \begin{matrix} if\ (\mathbf{T} \in [-R, +R]\ or\ \mathbf{T} \in [+R, -R])\ and\ Curvature = 1 \\ Otherwise \end{matrix} \tag{3}$$

where $Curvature = \begin{cases} 1 & if\ \eta \geq 0.15 \\ 0 & Otherwise \end{cases}$ and $\boldsymbol{R}$ is the quantitative measure of the change in $\boldsymbol{y}$ (Fig. 4).

The curvature detection algorithm was tested on various groups of hypotheses which demonstrated robust performance in terms of separating curved segments from linear ones (four of the groups given in Fig. 5)

2.3. Segment-manhole connection using expectation maximization algorithm

The connection of segments recognized from previous section with manholes was categorized as a classification problem where each segment was to be classified as a segment connected to two manholes. The EM algorithm was proposed for segment-manhole connection establishment which has been used in wide range of machine learning and data mining scenarios such as classification (Klautau, 2003; Sander et al., 1998), Image Processing (Huanhuan Chen, 2010) and unsupervised clustering (Bailey and Elkan, 1994; Buntine, 2002; Salojarvi et al., 2005). The inputs from sensor readings, estimated asset orientations from manholes, statutory records and segments recognized from previous step were used to identify suitable connections between segments and manholes. Combining the inputs created a data fusion scenario where hidden information from different sources was integrated and the probability of segment classification was updated until the algorithm converged (or the number of iterations reaches). Therefore, based on the probability distribution drawn from EM algorithm, the local maximum of segment-manhole connection provided an optimal solution that was then compared with prior information from statutory records. Using EM, each segment can be assigned to a pair of manhole

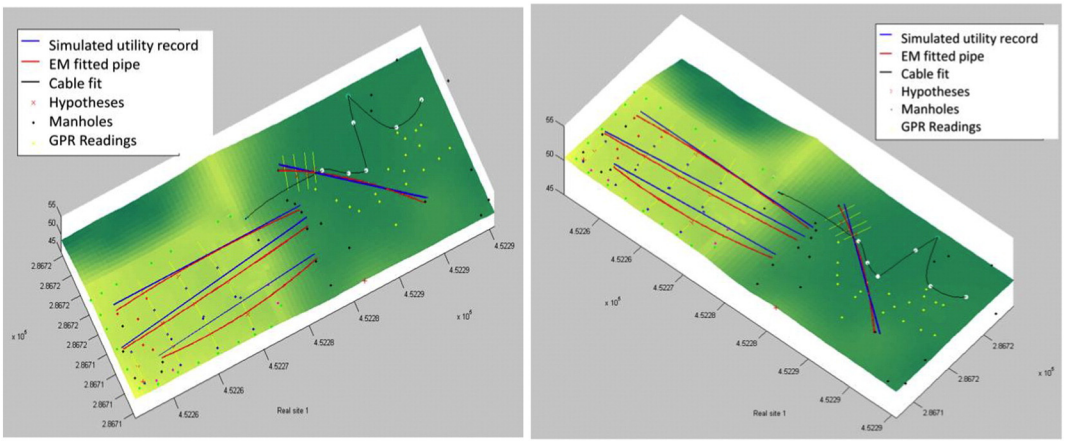

Fig. 10. Real Site 1. Blue lines represent simulated utility record based on manhole and sensor readings. Red lines represent segment-manhole connections established by the model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

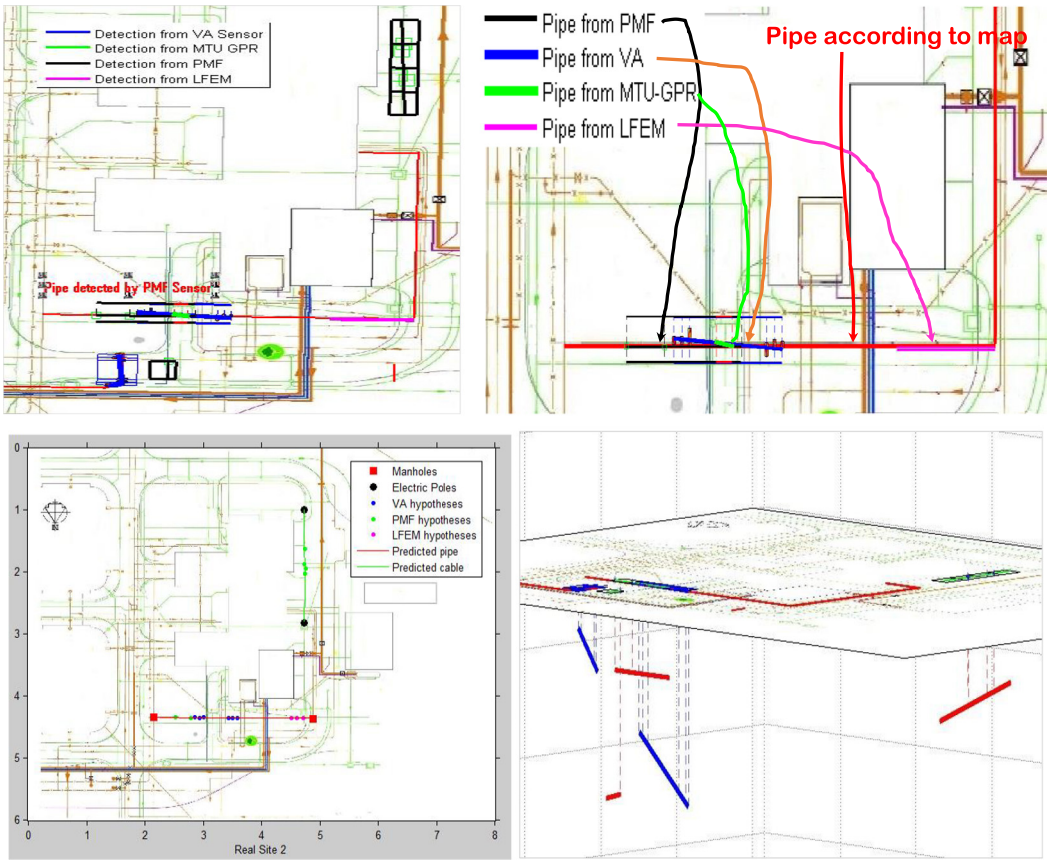

Fig. 11. Real survey site 2 **(a)** pipe detected by MTU sensors, **(b)** pipe detected by PMF sensor due to electric cable buried closely, **(c)** data from different sensors integrated together for segment-manhole connection, **(d)** 3D view of segments detected by Bayesian Mapping model.

connections given the above parameters. EM consists of two steps that are (1) Expectation and (2) Maximization. Expectation calculates the probability distribution of a target outcome given a set of parameters while the maximization step updates the distribution given the parameters which are updated from expectation at each iteration. The EM algorithm is explained as follows:

1) For a given hypothesis point h, get nearest segment S_j from statutory record. Get manhole locations and the depths. Initialise the prior for each manhole connection with P_i as uniform distribution. Initialise covariance matrix

$$\mathbf{\Lambda} = \; diag(\Delta d, \|\theta_j - \theta_k\|, \|\theta_k - \theta_{hk}\|, \|\theta_j - \theta_{hk}\|, \|Z_h - Z_k\|)$$

where $\Delta d = \sqrt{(x_{Sj} - x_k)^2 + (y_{Sj} - y_k)^2}$, θ_j = orientation of S_j, θ_k = orientation of segment of same type as of S_j from manhole k's cover, θ_{hk} = orientation from manhole k to h, z_h = depth of h from sensor, z_k = depth of the segment of same type as of S_j from manhole k's cover, $\mathbf{\Theta} = (K, X, Y, Z, x_h, y_h, z_h, S_j)^T$ for K manholes with $[X, Y, Z]$ as position and depth.

2) E-step: Likelihood for each manhole k connection

$$L\langle C_k|\mathbf{\Lambda}, \mathbf{\Theta}\rangle = \frac{1}{(2\pi|\mathbf{\Lambda}|)^{3/2}} \exp\left\{-\frac{1}{2}\left(\mathbf{C}(\mathbf{h}, \mathbf{k})^T \mathbf{\Lambda}^{-1} \mathbf{C}(\mathbf{h}, \mathbf{k})\right)\right\}$$

where $\mathbf{C}(\mathbf{h}, \mathbf{k}) = [\Delta d, \|\theta_j - \theta_k\|, \|\theta_k - \theta_{hk}\|, \|\theta_j - \theta_{hk}\|, \|z_h - z_k\|]$. The Maximum Likelihood Estimate (MLE) is then given by $\gamma_k = \frac{L\langle C_k|\mathbf{\Lambda}, \mathbf{\Theta}\rangle}{\sum_{k=1}^{K} L\langle C_k|\mathbf{\Lambda}, \mathbf{\Theta}\rangle}$

3) M-Step: Maximization of likelihood

$$\mathbf{\Lambda} = \; diag\left(\sqrt{\frac{\left(\sum_k \prod_{k=1}^{K} (\Delta d . L\langle C_k|\Lambda, \Theta\rangle)\right)^2}{\sum_{k=1}^{k} L\langle C_k|\Lambda, \Theta\rangle}}, \sqrt{\frac{\left(\sum_k \prod_{k=1}^{K} (\|\theta_j - \theta_k\| . L\langle C_k|\Lambda, \Theta\rangle)\right)^2}{\sum_{k=1}^{k} L\langle C_k|\Lambda, \Theta\rangle}}, \sqrt{\frac{\left(\sum_k \prod_{k=1}^{K} (\|\theta_k - \theta_{hk}\| . L\langle C_k|\Lambda, \Theta\rangle)\right)^2}{\sum_{k=1}^{k} L\langle C_k|\Lambda, \Theta\rangle}}, \sqrt{\frac{\left(\sum_k \prod_{k=1}^{K} (\|\theta_j - \theta_{hk}\| . L\langle C_k|\Lambda, \Theta\rangle)\right)^2}{\sum_{k=1}^{k} L\langle C_k|\Lambda, \Theta\rangle}}, \|z_h - z_k\|\right)$$

4) Repeat 2–3 until converged (log likelihood converges to 10^{-4})

5) Maximum A Posteriori (MAP) for connection of k with S_j is $C_{h \to k} = \arg \max_{k \in K}(\gamma_k)$

The MLE γ_k in step 2 provides best match between a manhole k and a segment S_j. The algorithm is repeated for connection from both ends of S_j to a pair of best matching manholes. At the test site, it is possible for sensors to contain noise (containing spatial location and direction error) as part of the hypotheses. A point h was considered noise when it's likelihood of manhole connections ($k_{\rightarrow}$ and $k_{\leftarrow}$) fell below a threshold $\alpha = \frac{\sum_{h \in O} P(L|h, k_{\rightarrow}, k_{\leftarrow})}{\frac{3}{2}O}$ where $P(L|h, k_{\rightarrow}, k_{\leftarrow}) \approx N(L, h, \sigma_o)$. Here, σ_o (=0.4) is the variance (meters) calculated from (x,y,z) of O. The noise from above set of hypotheses points was identified using

$$h = \begin{cases} noise & if\ P(L|h, k_{\rightarrow}, k_{\leftarrow}) < \alpha \\ h & otherwise \end{cases} \tag{4}$$

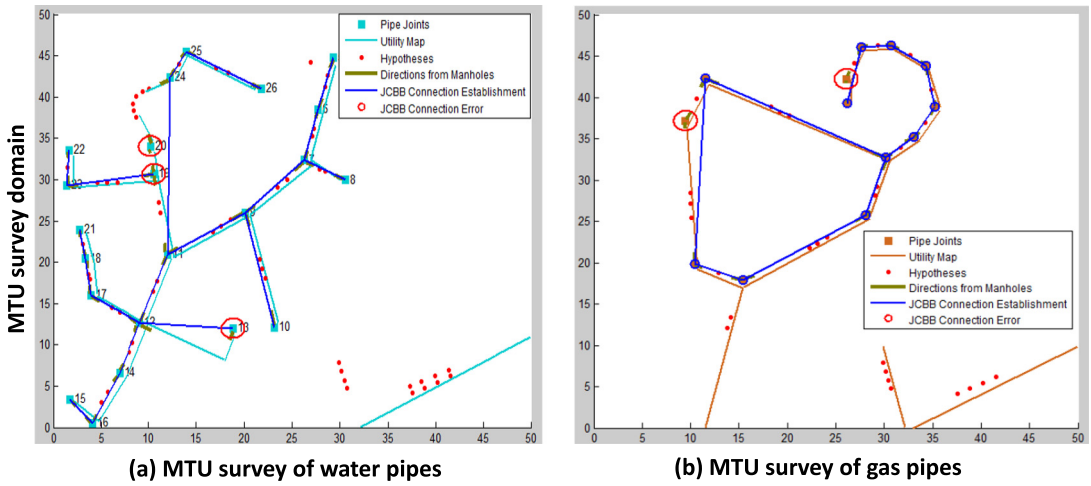

Fig. 12. (**a**) JCBB Connection Error in water pipes (**b**) JCBB Connection Error in gas pipes.

The clusters of segments for manhole connections from (4) consisted of the hypotheses points and the connected manhole positions which are: $\mathbf{C_j} = [k_{\rightarrow}\ \{h \in O : h \rightarrow S_j\}\ k_{\leftarrow}]$

3. Results and discussion

Dataset for model validation included five sites of simulated data (3 sets) as well as real data (2 sets). The model was developed and validated using MATLAB 2011(b) where the most probable maps were generated given 3 different sources of information: (i) sensor readings, (ii) manhole surveys, and (iii) the statutory records. In order to validate the model, hypotheses extraction, accurate segment identification (with noise removal) and segment-manhole connection were tested. Several variations in the parameters and data quality were considered in order to verify the robustness of the model. The hypotheses were extracted from raw sensor images using image segmentation techniques. A segment recognition algorithm was developed to identify linear and curved pipe/cable segments based on the direction of survey taken by MTU machine, and orientations estimated from hypotheses when they were combined together. Total of 4 initial use cases of hypotheses (Fig. 6) tested in order to evaluate the performance of segment recognition algorithm where the number of segments in each use case were known.

The simulated datasets (S1, S2, and S3) consisted of locations of manholes as well as pipes from ArcGIS (Esri, 2015) and simulated hypotheses (simulated readings from MTU sensors). The real sites (R1 and R2) consisted of measurements from MTU sensor readings and statutory records. For noise removal in simulated sets, white noise was added with the following: (1) Spatial noise of the locations of manholes and pipes was up to 2 × 2 m, (2) Noise in hypotheses locations was up to 0.4 m and (3) Noise in pipe directions was up to 8 degrees.

3.1. Bayesian mapping model for simulated data

The model was initially tested on 3 simulated datasets (S1, S2 and S3) with varying numbers of manholes, sensor data points, and segments from statutory records (Table 1). For simulated datasets, the hypotheses were manually generated and the initial maps were obtained from ArcGIS (Esri, 2015). Initial simulated datasets with initial maps are depicted in Fig. 7.

(a) Segment recognition: Segment recognition algorithm showed robust performance for all datasets and the noise was removed from groups of hypotheses in S1 and S2. The absence of statutory records did not affect the capability of the model to draw accurate segments due to the segment recognition algorithm. In addition, the segments were drawn in 3D due to the inclusion of depth information from manholes and sensor readings. However, the uncertainty increased in the absence of statutory records as they provided useful information on the number of buried assets and their directions. The segments drawn by automated segment recognition algorithm for simulated datasets are given in Fig. 8.

(b) Segment-manhole connection: The segment recognition algorithm provided advantages over (Chen and Cohn, 2011) in terms of basis for accurate connection with manholes and construction of 3D maps. The segments were used by EM algorithm for segment-manhole connection establishment. Using the parameters from different sources (Materials and Methods), it was observed that the algorithm successfully created connections for segments that led to the reconstruction of 3D map. Among the segments generated from step (a) a few segments had single end connecting to a manhole. Such situations were also tested for the validation of EM algorithm. The maps of simulated datasets were generated which are depicted in Fig. 9.

3.2. Bayesian mapping model for real data

Real sites (R_1 and R_2) with sensor readings and manhole surveys were tested for model validation. Table 1 provides information on the inputs to model for simulated sites as well as real sites. There were 18 sensor readings taken at R_1 which included readings for linear pipes

Table 2
Directional (**$D(\Theta)$**), Spatial (**$S(x,y)$**) and connection errors compared to JCBB. **EM\L** = without segment recognition algorithm, **EM\U** = without utility records.

Error	EM\L	EM\U	EM	JCBB	U	E(JCBB)	E(EM)
$S(x,y)_{S1}$	1.3514	1.7923	0.4132	3.1548	2.8	1	0
$D(\Theta)_{S1}$	3.3371	4.4451	1.2526	6.1112	4		
$S(x,y)_{S2}$	2.1334	3.5691	1.005	7.4324	2.5	0	0
$D(\Theta)_{S2}$	4.3351	6.4321	1.1129	10.9813	5		
$S(x,y)_{S3}$	3.1058	6.9471	2.0591	12.6976	6.5	5	0
$D(\Theta)_{S3}$	4.6131	8.8195	2.0031	16.9115	6		
$S(x,y)_{R1}$	4.3291	8.7795	1.6531	11.1125	5	1	0
$D(\Theta)_{R1}$	3.9973	6.9553	1.8533	8.8134	4		
$S(x,y)_{R2}$	2.4415	3.1101	1.0051	4.9518	3	0	0
$D(\Theta)_{R2}$	2.3705	3.1781	0.8109	4.1136	3		

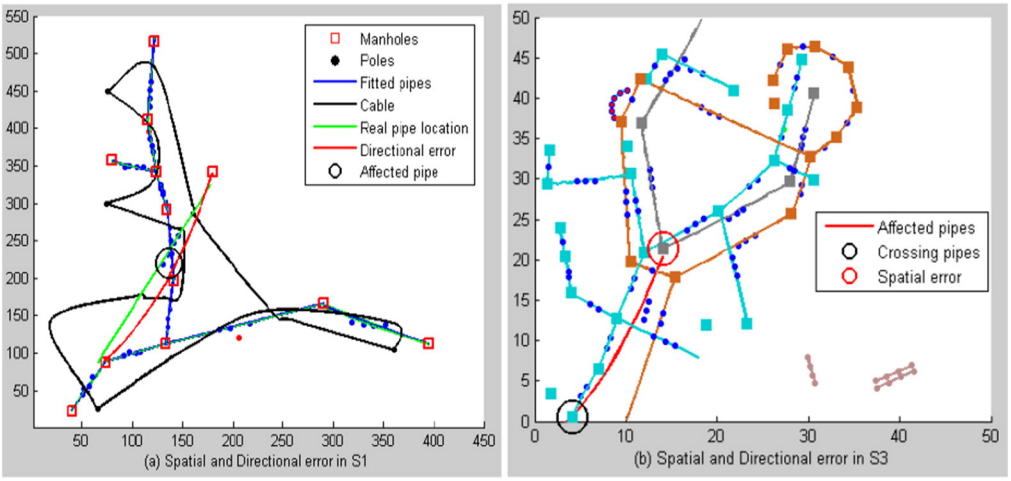

Fig. 13. Impact of spatial and directional error on survey sites S2 and S3.

and an electricity cable. Raw images from sensors were manipulated to extract hypotheses which were used for automatic segment recognition. There were no connection errors for R_1 using EM based segment-manhole connection algorithm for linear segments (See Fig. 10).

Real site R_2 consisted of straight linear pipes with an electric cable for which the sensor readings were taken (Fig. 11a). There was no noise detected in the sensor readings and segments were successfully connected with manholes. PMF sensor usually detects objects with electrical current such as electric cables. At R2, The PMF sensor detected a water pipe (rectangle in Fig. 11b) in the survey. This was due to the run of an electric cable in a close proximity of that water pipe which emits the electrical current.

3.3. Connection and spatial errors in mapping models

For manhole connection establishment process, the proposed algorithm was also compared with Joint Compatibility Branch & Bound (JCBB) which calculates the Mahalanobis distance between the observations and the predictions and accepts a connection if the Mahalanobis distance is smaller than a validation gate (Yangming Li et al., 2013). The segment-manhole connection problem was similar to the spatial data association problem in robotics (Bailey and Durrant-Whyte, 2006). Using the data fusion technique, noise removal and the iterative refinement of posterior probability using EM algorithm, the analysis showed better results compared to JCBB for both simulated sets as well as real sites. JCBB suffered in connecting segments in simulated scenarios with manholes when more than one segments were located in close proximity in the map. In tested datasets, JCBB also suffered from connection errors at the manhole locations where directional error exceeded five degrees. There were total five manhole connection errors recorded when JCBB was run on S3 as shown in Fig. 12 where Fig. 12a shows the connection errors for water pipes and Fig. 12b shows the connection errors for gas pipes.

The analysis with segment recognition algorithm significantly reduced the spatial errors (in meters) and directional errors (in degrees) when modeling for mapping the underground utilities. The mean directional error using EM algorithm was smaller compared to JCBB. The connection errors (i.e., false positives (number of manholes connections predicted for the wrong segment)) are given in Table 2 which are denoted as E(JCBB) for JCBB and E(EM) for EM algorithm.

A noticeable impact on the voxel prediction for each pipe was observed when this error rate was increased from 8 to 14 degrees. The increased directional error in S1 and S3 had significant effect with inaccurate segment-manhole connections in the maps. In addition, the spatial error resulted in variations in voxel classification as shown in Fig. 13a and b. The black circles show directional errors and the red circles show spatial error.

4. Conclusions

The traditional approaches for map generation given in the literature require information from statutory records and provide solutions for mostly 2D map generation. Additionally, the information from multiple sensors require extensive processing and statutory records should also be integrated that provide prior knowledge about buried utilities. Recent analyses/studies on the abilities to locate underground utilities stressed the needs of a system capable of generating real time maps with the help of heterogeneous information. However, current techniques are limited to either generating 2D maps of linear segments or unable to detect different types of utilities. This paper addresses these problems by proposing a Bayesian mapping model by implementing various machine learning techniques for real time 3D map (re)construction. The segment recognition algorithm is robust in identifying groups of hypotheses forming linear/curved segments that are helpful in establishing connections with manholes. In order to improve utility classification and refined map generation, noise removal facilities were embedded in the system that improved performance in distinguishing between hypotheses and noise. The Bayesian Mapping model is aimed to overcome critical issues related to efficient and real-time location of buried assets that could provide valuable underground information and be cost effective. Further online analysis will also validate model performance at higher levels and its ability to generate refined maps at real time.

Contributions

The research and development activities were conducted mainly at the University of Leeds, under the supervision of Professor Anthony G. Cohn. First author conducted the analysis, prepared results and the manuscript under the supervision of Prof. Anthony Cohn, second author conducted major revisions and improvements in the paper, while other authors at partner institutions contributed to the

development of MTU sensors and providing raw data from Wigan and Blagdon test sites.

Conflicts of interest

The authors do not have any conflict of interest.

Acknowledgements

This research is supported by EPSRC grants (EP/F06585X/1 and EP/K021699/1): Mapping the Underworld: Multisensory Device Creation, Assessment, Protocols (http://www.mappingtheunderworld.ac.uk). We are thankful to the project management, especially, Mark Hamilton, for all site visits, collaboration and management of the efforts from all project partners.

Appendix A. Appendix

For curved segments, the three-point circle fitting was applied as illustrated in Fig. A-1.

A. Most probable segment estimation

In order to refine the output of linear segment fitting without significant impact on model performance, the least squares fitting was performed on the above sets of hypotheses and manhole connections. Even though the data from multiple sources was integrated for Bayesian mapping, the sensor readings may contain small non-negligible noise which may also affect the performance of least squares linear fitting algorithm. Fig. A-2 shows a simple scenario where the point containing noise in its location causes an increased angular difference with the orientation from connecting manhole. Also, the segment approximation fails to find the most probable linear segment without crossing any other segment in close proximity.

It is critical to develop an approach for above mentioned scenarios since it is common that the street surveys contain more than one pipes buried in close proximity and the chances of having the overlaps between those segments are higher. In order to overcome such challenge, a probability distribution over the space of three possible lines L_i ($i = 1,2,3$) was created and described by discrete random variables. L_i is the i^{th} line of voxels given the specific attributes which are the locations of the points of clusters defined in previous section. We need to find

$$\arg\max_{l} \prod_{l=1}^{g_l} (P(y_l|L_1, L_2, L_3)) \tag{A1}$$

and

$$\arg\max_{l} \prod_{l=1}^{g_l} (P(z_l|L_1, L_2, L_3)) \tag{A2}$$

where $l \in \boldsymbol{g_l}$ is the voxel chosen from a set $\boldsymbol{g_l}$ of voxels fitted for the quantized line. The probability distribution for three observed possible 3D positions of each line L_i were created and combined given x, y and z information. For each voxel of a line L_i, two equal spaced perpendicular line segments $\not{p}$ were drawn along the y-axis for Eq. (**A1**) and z-axis for Eq. (**A2**). This set of linear segments is denoted by $\mathcal{G}$. For the calculation of the most probable voxel at each quantised position of each line L_i, we use two terminologies. For Eq. (**A1**), the pixel p_i^l of L_i is the (x_l, y_l) and for Eq. (**A2**), the pixel p_i^l of L_i is the (y_l, z_l). Therefore Eq. (**A1**) is used for determining the most probable y^{th} position of the voxel and Eq. (**A2**) is used to determine the z^{th} position of the voxel in the map. For all k voxels defined by $\boldsymbol{g_l}$, the joint probability distributions are given by

$$P(y_l|L_1, L_2, L_3) = \prod_{i=1}^{L} P(y_l|L_i) \; for \; L = 3$$

$$P(z_l|L_1, L_2, L_3) = \prod_{i=1}^{L} P(z_l|L_i) \; for \; L = 3$$

where

$$P(y_l|L_i) = \sum_{l=1}^{g_l} \sum_{y \in G} \frac{1}{\|\nabla y\|} \exp\left\{ -\frac{\|y - p_i^l\|^2}{\|g_p\|^2} \right\} \tag{A3}$$

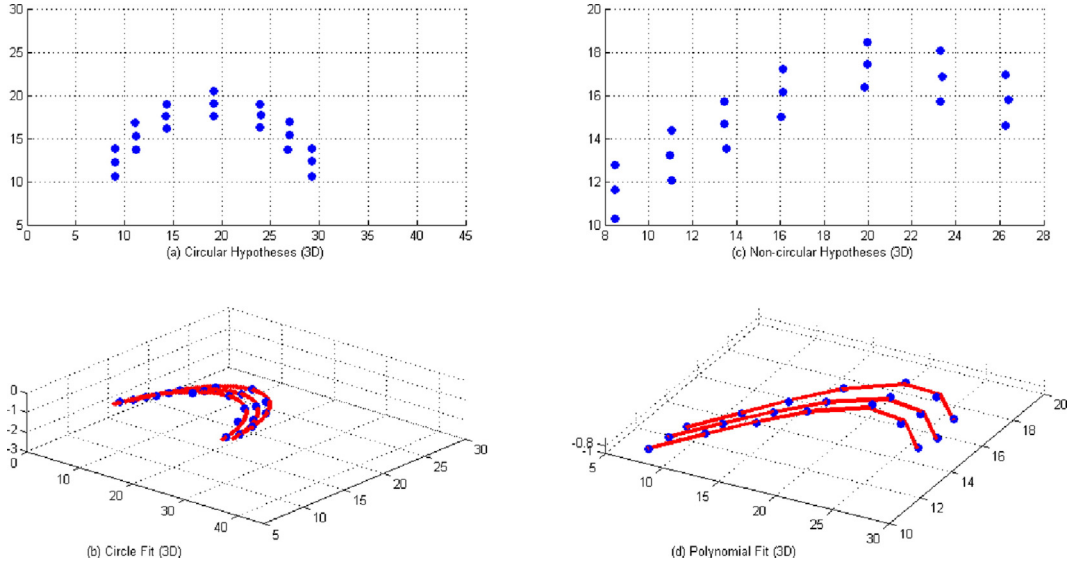

Fig. A-1. **(a)** Circular hypotheses, **(b)** Circle fit for circular points, **(c)** Non-circular Hypotheses, **(d)** Polynomial fit for non-circular points.

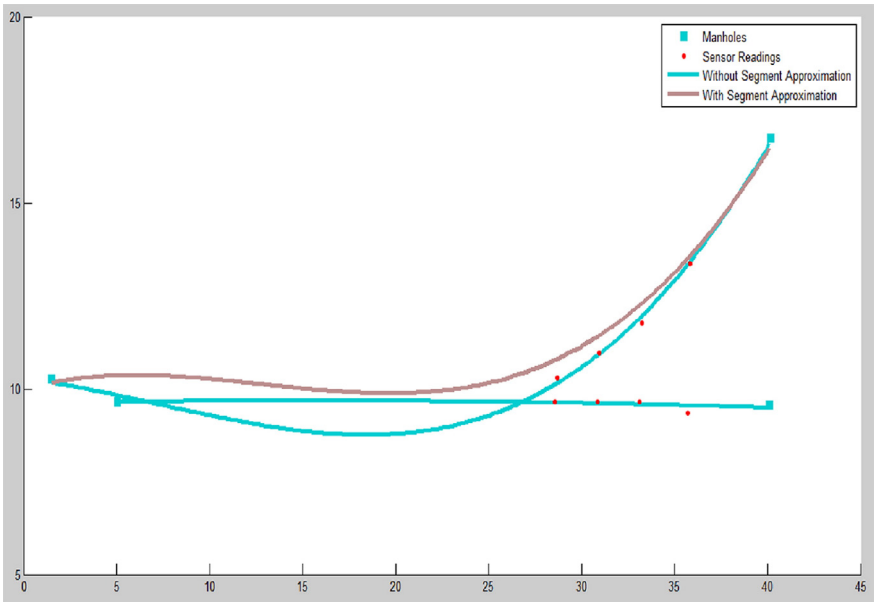

Fig. A-2. Pipe segment fitted from hypotheses crossing another closed segment in the map.

p_i^l is the l^{th} pixel (x_l, y_l) of L_i

$$P(z_l|L_i) = \sum_{l=1}^{g_l} \sum_{z \in G} \frac{1}{\|\nabla z\|} \exp\left\{ -\frac{\|z - p_i^l\|^2}{\|g_p\|^2} \right\} \quad \text{(A4)}$$

p_i^l is the l^{th} pixel (y_l, z_l) of L_i

Calculating $P(y_l|L_i)$ and $P(z_l|L_i)$ provide the probabilities $P(y_l|L_1,L_2,L_3)$ and $P(z_l|L_1,L_2,L_3)$ respectively which are used to calculate the most probable voxel position for the map using

$$P(L_y|L_1, L_2, L_3) = \arg \max_l \prod_{l=1}^{g_l} (P(y_l|L_1, L_2, L_3)) \quad \text{(A5)}$$

and

$$P(L_z|L_1, L_2, L_3) = \arg \max_l \prod_{l=1}^{g_l} (P(z_l|L_1, L_2, L_3)) \quad \text{(A6)}$$

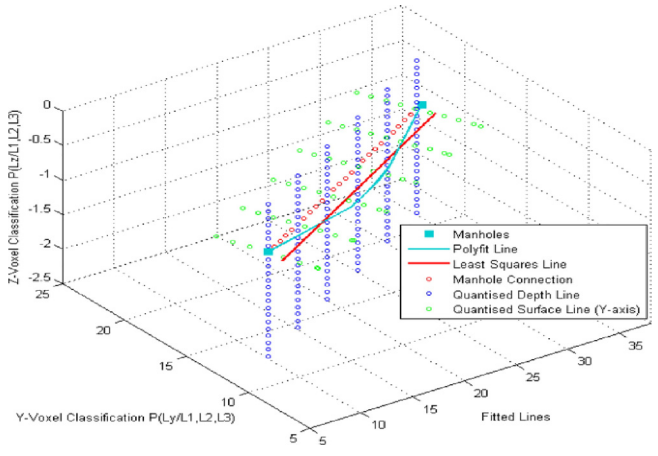

Fig. A-3. Perpendicular quantised linear points created from the point and the linear segment.

An example of this is shown in Fig. A-3 where three lines (L_1,L_2,L_3) are drawn from three techniques and the quantized lines on y-axis and z-axis are drawn to calculate Eqs. (A5) and (A6).

At each step in the map, the voxel with the highest probability given three lines as input was taken as the most probable voxel. The set of all voxels at the end created a line in 3D space which was considered as the most probable linear segment approximation. This algorithm was applied for the approximation of each linear segment in the performance evaluation of the algorithm.

B. Cardinal spline fitting

In addition to the linear and curved pipe segments which may include sewer pipes, gas pipes and water pipes, the street maps also included electric cables buried underground which need to be approximated. The difference of the cables from the other pipe segments is their non-linear behavior. The cable can be buried in an arbitrary order without satisfying linearity condition for which line fitting or curve approximation may represent inappropriate solutions. The Passive Magnetic Field (PMF) sensors report the sequence of the survey locations when performing street survey. For fitting cables and non-linear segment, Cardinal Spline algorithm was implemented which was capable for fitting the lines to points in both 2D and 3D. The detailed explanation of Cardinal Spline algorithm is provided elsewhere (Ali Khan and Sarfraz, 2011).

C. Probabilistic voxel classification

Errors in the linearity of segment lines vary depending on the noise added in sensor and manhole readings. In addition, the presence of multiple closely-buried assets is challenging when recognizing the number of segments. In order to address these challenges, linear approximation algorithm was useful as it integrated information from multiple sources and provided most probable estimation of voxel classification. Bayesian probabilistic voxel classification showed improved results in terms of fitted lines when compared to least squares fitting.

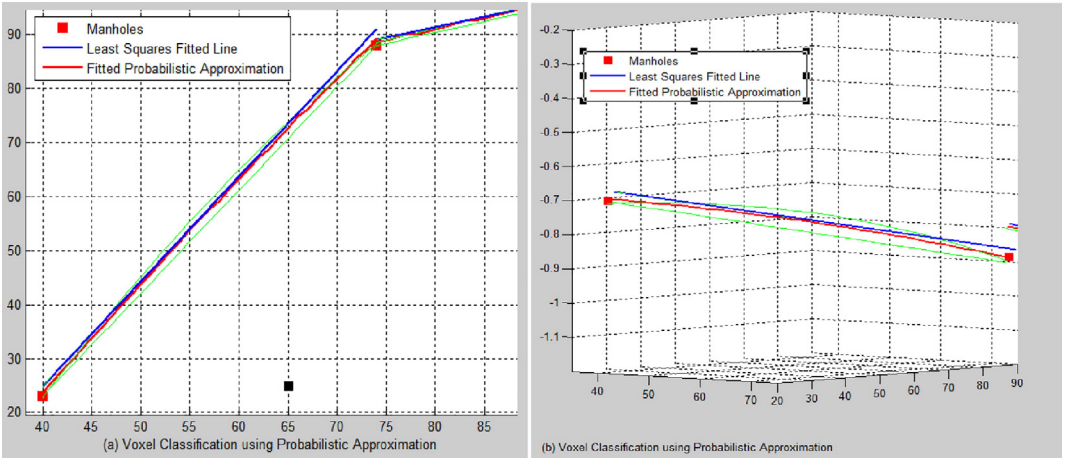

Fig. C1. Lines regressed for two pipes with and without using probabilistic approximation algorithm

References

Aleardi, M., Ciabarri, F., Mazzotti, A., 2017. Probabilistic estimation of reservoir properties by means of wide-angle AVA inversion and a petrophysical reformulation of the Zoeppritz equations. J. Appl. Geophys. 147 (Supplement C):28–41 Available at:. http://www.sciencedirect.com/science/article/pii/S0926985117301192.

Ali Khan, M., Sarfraz, M., 2011. Motion tweening for skeletal animation by cardinal spline. In: Abd Manaf, A., Sahibuddin, S., Ahmad, R., Mohd Daud, S., El-Qawasmeh, E. (Eds.), Informatics Engineering and Information Science. Springer, Berlin Heidelberg.

Al-Nuaimy, W., et al., 2000. Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition. J. Appl. Geophys. 43 (2):157–165 Available at. http://www.sciencedirect.com/science/article/pii/S0926985199000555.

Ashdown 2000. Mains Location Equipment – A State of the Art Review and Future Research Needs. In: UKWIR (Ed.). (ISBN: 1 84057 233 7).

Bailey, T., Durrant-Whyte, H., 2006. Simultaneous localization and mapping (SLAM): part II. IEEE Robotics & Automation Magazine 13 (3), 108–117.

Bailey, T.L., Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings International Conference on Intelligent Systems for Molecular Biology. 2, pp. 28–36.

Bar-Shalom, Y., 1987. Tracking and Data Association. Academic Press Professional, Inc., San Diego, CA, USA.

Bhalerao, Abhir, Wilson, R., 2001. A Fourier Approach to 3D Local Feature Estimation from Volume Data (2001). British Machine Vision Conference. pp. 461–470.

Buntine, W., 2002. Variational Extensions to EM and Multinomial PCA. In: Elomaa, T., Mannila, H., Toivonen, H. (Eds.), Machine Learning: ECML 2002. Springer Berlin Heidelberg.

Burtwell, M., E. A., 2004. Locating underground plant and equipment - proposals for a research programme. UKWIR (ISBN: 1-84057).

Chen, Huanhuan, A. G. C, 2010. Probabilistic robust hyperbola mixture model for interpreting ground penetrating radar data. The 2010 International Joint Conference on Neural Networks (IJCNN). IEEE, Barcelona.

Chen, H., Cohn, A.G., 2011. Buried utility pipeline mapping based on multiple spatial data sources: A Bayesian data fusion approach. IJCAI Retrieved from. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-402.

Chen, T.W., Wang, Q., 2010. 3D line segment detection algorithm for large-scale scenes. 2010 International Conference on Audio, Language and Image Processing: pp. 1056–1061 https://doi.org/10.1109/ICALIP.2010.5685113.

Christensen, C.W., et al., 2015. Combining airborne electromagnetic and geotechnical data for automated depth to bedrock tracking. J. Appl. Geophys. 119 (Supplement C):178–191 Available at:. http://www.sciencedirect.com/science/article/pii/S0926985115001585.

Delicado, P., Smrekar, M., 2007. Mixture of nonlinear models: a Bayesian fit for principal curves. 2007 International Joint Conference on Neural Networks:pp. 195–200 https://doi.org/10.1109/IJCNN.2007.4370954.

Dettmer, J., et al., 2014. Trans-dimensional finite-fault inversion. Geophys. J. Int. 199 (2): 735–751 Available at. https://doi.org/10.1093/gji/ggu280.

Do, C.B., Batzoglou, S., 2008. What is the expectation maximization algorithm? Nat. Biotechnol. 26:897. https://doi.org/10.1038/nbt1406.

Dong-Min, W., Dong-Chul, P., 2009. 3D line segment detection based on disparity data of area-based stereo. Intelligent systems, 2009. GCIS '09. WRI Global Congress on, 19–21 May 2009, pp. 219–223.

Dutta, R., Cohn, A.G., Muggleton, J.M., 2013. 3D mapping of buried underworld infrastructure using dynamic Bayesian network based multi-sensory image data fusion. J. Appl. Geophys. 92:8–19 Available at:. http://www.sciencedirect.com/science/article/pii/S0926985113000311.

Esri, 2015. Put Your Maps to Work with ArcGIS, the Mapping Platform for Your Organization [Online]. Esri Available:. https://www.arcgis.com.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. AAAI Press, pp. 226–231.

Fernández-Martínez, J.L., et al., 2013. From Bayes to Tarantola: new insights to understand uncertainty in inverse problems. J. Appl. Geophys. 98 (Supplement C), 62–72.

Fiannacca, P., et al., 2017. IG-Mapper: A new ArcGIS® toolbox for the geostatistics-based automated geochemical mapping of igneous rocks. Chem. Geol. 470 (Supplement C), 75–92.

Friedman, J., Popescu, B., 2004. Gradient Directed Regularization for Linear Regression and Classification (Retrieved from citeulike-article-id:3142778).

Gonzalez, R.C., 2008. Digital Image Processing. Third Ed.

Guo, R., et al., 2011. Non-linearity in Bayesian 1-D magnetotelluric inversion. Geophys. J. Int. 185 (2), 663–675.

Ji, Y., et al., 2016. Frequency-domain sparse Bayesian learning inversion of AVA data for elastic parameters reflectivities. J. Appl. Geophys. 133 (Supplement C), 1–8.

Klautau, Aldebaro, N. J. A. A. O, 2003. Discriminative Gaussian mixture models: a comparison with Kernel classifiers. In: Mishra, T.F.A.N. (Ed.), Proceedings of the 20th International Conference on Machine Learning (ICML-03).

Krishna, A.S., Rao, G.S., 2013. Contrast enhancement techniques using histogram equalization methods on color images with poor lightning. Int. J. Comput. Sci. Eng. Appl. 15–24.

Mcmahon, W., Burtwell, M.H., Evans, M., 2005. Minimising street works disruption: the real costs of street works to the utility industry and society. UKWIR.

Muggleton, J.M., Brennan, M.J., Gao, Y., 2011. Determining the location of buried plastic water pipes from measurements of ground surface vibration. J. Appl. Geophys. 75 (1), 54–61.

Neira, J., Tardos, J.D., 2001. Data association in stochastic mapping using the joint compatibility test. IEEE Trans. Robot. Autom. 890–897.

Oldenborger, G.A., et al., 2016. Bedrock mapping of buried valley networks using seismic reflection and airborne electromagnetic data. Journal of Applied Geophysics 128 (Supplement C), 191–201.

Otsu, Nobuyuki, 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern, SMC-9 62–66.

Ren, H., et al., 2017. Bayesian inversion of seismic and electromagnetic data for marine gas reservoir characterization using multi-chain Markov chain Monte Carlo sampling. J. Appl. Geophys. 147 (Supplement C), 68–80.

Ristić, A., et al., 2017. Point coordinates extraction from localized hyperbolic reflections in GPR data. J. Appl. Geophys. 144 (Supplement C), 1–17.

Royal Acd, R. C., Atkins Pr, Brennan Mj, Chapman Dn, Chen H, Cohn Ag, Curioni G, Foo Ky, Goddard K, Hao T, Lewin Pl, Metje N, Muggleton Jm, Naji A, Pennock Sr, Redfern Ma, Saul Aj, Swingler Sg and Wang P 2010. Mapping the Underworld: Location, Mapping and Positioning Without Excavation. In: Control, C. (Ed.). (Middle East, Singapore).

Royal, A.C.D., Atkins, P.R., Brennan, M.J., Chapman, D.N., Chen, H., Cohn, A.G., Foo, K.Y., Goddard, K.F., Hayes, R., Hao, T., Lewin, P.L., Metje, N., Muggleton, J.M., Naji, A., Orlando, G., Pennock, S.R., Redfern, M.A., Saul, A.J., Swingler, S.G., Wang, P., Rogers, C.D.F., 2011. Site assessment of multiple-sensor approaches for buried utility detection. Int. J. Geophys. 2011, 19.

Salojarvi, J., Puolamaki, K., Kaski, S., 2005. Expectation maximization algorithms for conditional likelihoods. Proceedings of the 22nd International Conference on Machine Learning. ACM, Bonn, Germany.

Sander, J., Ester, M., Kriegel, H.-P., Xu, X., 1998. Density-based clustering in spatial databases: the algorithm GDBSCAN and its applications. Data Min. Knowl. Discov. 2, 169–194.

Sanquer, M., Chatelain, F., El-Guedri, M., Martin, N., 2011. A reversible jump MCMC algorithm for Bayesian curve fitting by using smooth transition regression models. Acoustics, Speech and Signal Processing (ICASSP). 2011 IEEE International Conference on, 22–27 May 2011, pp. 3960–3963.

Thomas, A.M., et al., 2009. Stakeholder needs for ground penetrating radar utility location. J. Appl. Geophys. 67 (4), 345–351.
Underworld, M.T., 2011. Mapping the Underworld [Online]. Available:. http://www.mappingtheunderworld.ac.uk, Accessed date: 13 September 2015.
Wang, Y., Lu, W., 2016. Discontinuity enhancement based on time-variant seismic image deblurring. J. Appl. Geophys. 135 (Supplement C), 155–162.
Ward, G., Hastie, T., Barry, S., Elith, J., Leathwick, J.R., 2009. Presence-only data and the em algorithm. Biometrics 65 (2):554–563. https://doi.org/10.1111/j.1541-0420.2008.01116.x.
Werman, M., Keren, D., 1999. A novel Bayesian method for fitting parametric and non-parametric models to noisy data. Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on 1999 Vol. 2, pp. 1–558.
Yangming Li, S.L., Song, Quanjun, Liu, Hai, Meng, M.Q.-H., 2013. Fast and robust data association using posterior based approximate joint compatibility test. IEEE Transactions on Industrial Informatics, pp. 331–339.